TECHNICAL APPENDIX: CHAPTER VII

ECONOMETRIC MODEL, STATISTICAL RESULTS AND FORECASTING METHODOLOGY

Econometric Model Specification

To take full advantage of the available data and to allow for estimation of the effect that acres has on wilderness visitation, it was advantageous to pool the time series data over the four U.S. Census/RPA regions. The time series nature of the data and the pooling of time series and cross section raise several econometric issues. As with all time series data, the possibility of autocorrelation is a serious issue. Autocorrelation implies a serial correlation of error terms. Specifically, that a disturbance or pertabation in one period does have an effect on future periods. Such a correlation violates one of the assumptions of OLD regression. While the regression coefficients themselves are unbiased and consistent, their variances are biased making significance tests misleading (Kmenta, 1986:311). Preliminary OLD analysis with this data resulted in Durbin-Watson statistics strongly suggesting autocorrelation. As part of the statistical analysis, we corrected for first-order serial correlation using a first-order autoregressive correction. This procedure incorporates the residual from the past observation into the regression model for the current observation. Equation (1) provides the AR(1) corrected equation that is estimated. Take particular note of the error term:

$$\text{(A1) } (RVD/POP)it = \alpha + \beta Xit + (\rho\epsilon_{it-1} + \eta_{it})$$

where α and β are intercept and slope coefficients, respectively, to be estimated. Subscript I refers to Region I and t to time period t. ρ is the first-order serial correlation coefficient which is multiplied by the previous periods error term and then added to the unconditional error, η_{it}.

When pooling cross-section and time series data, it is useful to take advantage of the information that the data is really several blocks of related data. To enhance the variability in acres pooling data across regions is desirable. However, visitation in 1990 from the Pacific Coast may be different than visitation in 1990 from the East Coast due to unobserved factors not included as variables in the regression. With our use of four Census/RPA regions we essentially have four panels or groups of data. Running simple OLD regression does not take advantage of the fact that blocks of the observations are related. There are two approaches to incorporating the panel nature of this data into the estimation: (a) random effects; (b) fixed effects. As Greene (1990:485-6) notes, when the analyst has a census or 100 percent of the population data, the fixed effects model is likely to be more appropriate. In particular, we might expect each region to have a parametric shift in the regression function that relate to factors specific to that region. These show up as region specific constants. The fixed effects regression model in Limdep (Greene, 1995) uses a Likelihood Ratio and F-test to see if these region specific constants improve the fit over a single constant as would be estimated with OLD. Thus the fixed effects model is a generalization of OLD. A further advantage of the fixed effects model is that it avoids the assumption of the random effects model which assumes the individual effects are uncorrelated with the other regressors (Greene, 1990:495).

Incorporating fixed effects constant terms into the AR(1) corrected model in equation (A1) yields equation (A2):

$$\text{(A2)}\ \ln(RVD/POP)it = \alpha Di + \beta Xit + (\rho\epsilon_{it-1} + \eta_{it})$$

where Di are the regional constants reflecting the fixed effects and I= 1, 2, 3, 4 reflecting northeastern, southeastern, Rocky and Pacific Coast in the NPS model and I =1, 2 representing eastern U.S. and western U.S. in the USFS model. The difference in region specific constants between USFS and NPS resulted from comparative analysis of the same fixed effects structure for the two agencies. The fixed effect regression using the four Census/RPA regions resulted in the NPS model having a higher adjusted R square (.77 vs .61) as well as having substantially higher t-statistics on the regional dummies. Finally, the Likelihood Ratio tests and F-tests suggested the four Census/RPA regions were significant determinants of NPS visitor use as compared to the simpler east-west distinction. In addition, the NPS east-west fixed effects model substantially underestimated use in three of the regions. Detailed statistical results are presented in Appendix 1. For the USFS, the east-west region constants had t-statistics that were over four times higher than the Census/RPA regions (t-statistics of nearly 8 for the east-west region constants versus t-statistics of about 2 for the Census/RPA regions).

Estimation of separate models with slight differences in fixed effects specification of the NPS and USFS equations is consistent with results of Chow tests (Kmenta, 1986), which rejected coefficient equality for the two agencies Wilderness Areas at the .01 level. For both the USFS and NPS, we started with a full specification including all of the candidate independent variables listed above. Variables which were consistently insignificant (e.g., unemployment rate) were dropped from the final model. Retention of insignificant variables increases the variance of the regression.

Statistical Results

Table A-1 presents the results of the fixed effects analysis for Wilderness Areas administered by the USFS. Log of acres is significant at the .01 level while log of disposable per capita income is significant at the .02 level. Because of the double log specification, the coefficient on acres can be interpreted as an elasticity. Thus a 10 percent increase in Wilderness acres results in a 9 percent increase in recreation visitor days. This variable will allow us to predict changes in visitation with additions to the NWPS.

When the full model with age was run, age was consistently insignificant and so was dropped. The adjusted R square is quite high at .97 and the very large F statistic is significant at the .01 level. Each of the Region fixed effect constants are significant at the .01 level. While these constants may look only marginally different, the reader should recall that log of RVD's per capita is the dependent variable and hence even changes in the first decimal has a marked effect on the retransformed estimate of total RVDS.

As shown in Table A-2, the full fixed effect model with X variables and the group (region) effects out performs the classic multiple regression model (#3 X variables only) according to the likelihood ratio test Chi-Square statistic of 7.785 and the F test of 7.74, both of which are significant at the .01 level. Thus the fixed effects model will be used for forecasting future recreation use of wilderness.

Table A-3 presents the results of the fixed effects analysis for Wilderness Areas administered by the National Park Service. Here, log of acres, log of disposable per capita income and year are statistically significant at conventional levels. The adjusted R square of .77, while lower than the USDA Forest Service regression, is still quite good. The F statistic indicates the overall regression is significant at the .01 level. Each of the four Census/RPA Region fixed effect constants are significant at the .01 level.

As shown in Table A-4 the full fixed effect model with X variables and group (region) effects outperforms the classic multiple regression model (#3 X variables only) according to the likelihood ratio test Chi-square statistic of 28.465 and F test of 10.387, both of which are significant at the .01 level.

Forecasting

Forecasting procedure-- The forecasting procedure with the first order autoregressive correction is similar to regular forecasting except that the autoregressive correlation coefficient is multiplied by the previous periods residual. Specifically:

$$\text{(A3) est } (\ln(RVD/POP)it) = \alpha Di + \beta Xit + (\rho\epsilon_{it-1})$$

Each region's specific fixed effects constant (Di) is also used in forecasting that region's visitation.

As might be expected with such a high R square, the error in forecasting Forest Service Wilderness use across all regions and all years was just a few percent. Of course the annual average error by region was higher, with the average for the Northeast, Southeast, Rocky Mountain and Pacific Coast being about

3.3 percent, 1.5 percent, .7 percent and 4.8 percent, respectively. The National Park Service regression model also had a good overall prediction with an average error rate of just 2.6 percent across all regions. The average annual percent error by Census region were higher than this in the Southeast at 5.6 percent and 13 percent and 14 percent in the Rocky Mountain and Pacific Coast, respectively. To be conservative in our forecasts we have not performed the adjustment to the log retransformation suggested by Stynes, et al. (1986). Therefore, the forecasts represent the median, rather than mean visits.

Source of input values-- The accuracy of future forecasts of visits is equally dependent upon the future estimates of the independent variables as it is upon the coefficient estimates themselves. As part of the RPA Assessment process the USDA Forest Service commissioned the USDA Economic Research Service's Macroeconomics Team to estimate several future demographic variables including disposable personal income and unemployment rates (Torgeson, 1996). State level population forecasts were developed from U.S. Census projections and Bureau of Economic Analysis data by Dr. Linda Langner of the RPA staff.

This insured consistency in forecasting assumptions making possible aggregation of various forecasts. Generally the forecasted future values of the input variables are in line with recent trends. For example, disposable income is projected to grow at 2 percent a year, which is well below the historic time period but consistent with the experience of the last 6 years. One of the biggest unknowns is future wilderness acreages. Our initial forecast started with current quantity of wilderness as a baseline. Then we added agency's recommended wilderness acreage for the most likely future supply scenario to forecast recreation use. Finally, if information on total roadless acreage was available this was used as an upper limit on potential Wilderness supply for forecasting future recreation use.

Technical Appendix TableVII.1–Results of Fixed Effects Autoregressive Model for USDA Forest Service Wilderness

Variable	Coefficient	Standard Error	t-ratio	Prob	Mean
LACRES	0.90179	0.39617E-01	22.763	0.00	14.42
LPCINC	0.52445	0.22848	2.295	0.02	9.533
Group	Estimated Group Size	Fixed Effects Coefficient	Standard Error	t-ratio	Prob
1	57	-14.95145	1.87212	7.986	0.00
2	57	-14.60267	1.83469	7.979	0.00
Adjusted R square= .973 F Statistic= 1358.55 ρ=.8259					

Technical Appendix Table VII.2–Test statistics for the Classical Model versus Fixed Effects USDA Forest Service Wilderness Use Equation

Model	Log-Likelihood	Sum of Squares	R-squared
(1) Constant term only	-233.81048	0.403534E+03	0.0000000
(2) Group effects only	-160.50807	0.111525E+03	0.7236279
(3) X-variables only	-30.28344	0.113545E+02	0.9718624
(4) X and group effects	-26.39101	0.106050E+02	0.9737198

Hypothesis Tests							
	Likelihood Ratio Test			F Tests			
	Chi-quared	d.f.	Prob value	F	num.	denom.	Prob value
(2) vs (1)	146.605	1	0.00000	293.251	1	111	0.00000
(3) vs (1)	407.054	2	0.00000	1916.952	2	111	0.00000
(4) vs (1)	414.839	3	0.00000	1358.552	3	111	0.00000
(4) vs (2)	268.234	2	0.00000	523.399	2	111	0.00000
(4) vs (3)	7.785	1	0.00527	7.774	1	111	0.00624

Technical Appendix Table VII.3–Results of Fixed Effects Autoregressive Model for National Park Service Wilderness

Variable	Coefficient	Standard Error	t-ratio	Prob	Mean of X
LACRES	0.57074	0.14864	3.840	0.00	13.26
LPCINC	-5.8779	2.3540	-2.497	0.01	9.622
YEAR	0.12474	0.41994E-01	2.970	0.00	1984

Estimated Fixed Effects				
Region	Sample Size	Coefficient	t-ratio	Prob
1	18	-198.77400	3.18	0.00
2	18	-199.46114	3.19	0.00
3	23	-199.69602	3.20	0.00
4	22	-196.86188	3.18	0.00
Adjusted R square = .77 F=46.96 ρ=.499				

Technical Appendix Table VII.4–Test Statistics for the Classical Model versus Fixed Effects of National Park Service Wilderness Area

Model	Log-Likelihood	Sum of Squares	R-squared
(1) Constant term only	-146.60122	0.177036E+03	0.0000000
(2) Group effects only	-116.17783	0.835262E+02	0.5281975
(3) X-variables only	-97.23790	0.523268E+02	0.7044294
(4) X and group effects	-83.00525	0.368216E+02	0.7920112

Hypothesis Tests							
	Likelihood Ratio Test			F Tests			
	Chi-quared	d.f.	Prob value	F	num.	denom.	Prob value
(2) vs (1)	60.847	3	0.00000	28.735	3	76	0.00000
(3) vs (1)	98.727	3	0.00000	61.171	3	77	0.00000
(4) vs (1)	127.192	6	0.00000	46.965	6	75	0.00000
(4) vs (2)	66.345	3	0.00000	31.287	3	75	0.00000
(4) vs (3)	28.465	3	0.00000	10.387	3	75	0.00001